

ALGORITMA ASOSIASI DENGAN ALGORITMA APRIORI UNTUK ANALISA DATA PENJUALAN

Mohammad Badrul

Program Studi Sistem Informasi

Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri Jakarta

Jl. Damai No. 8 Warung Jati Barat Margasatwa Jakarta Selatan.

Telp. (021) 78839513 Fax. (021) 78839421

mohammad.mbl@nusamandiri.ac.id

Abstract — Each company or organization that wants to survive needs to determine the right business strategy. Data sales of products made by the company will eventually generate mountains of data. So it is very unfortunate if not analyzed back. products offered variety with a wide range of products, and sometimes the brand influence people to buy the product, to know the products with the highest sales and the linkage of products with each lainnya needed one existing algorithms in data mining algorithms are algorithms priori to be informed, and with the help of Tanagra applications, products which appear simultaneously knowable. Priori algorithms including the type of association rules in data mining. One association analysis phase which attracted the attention of many researchers to produce efficient algorithms is the analysis of patterns of high frequency (frequent pattern mining). Important or not an association can be identified by the two benchmarks, namely: support and confidence. Support (support value) is the percentage of the combination of these items in the database, while confidence (value certainty) is a strong relationship between the items in the rules of association. Priori algorithm can be helpful for the development of marketing strategies.

Intisari — Setiap perusahaan maupun organisasi yang ingin tetap bertahan perlu untuk menentukan strategi dalam bisnis yang tepat. Data penjualan produk yang dilakukan oleh perusahaan lambat laun akan menghasilkan tumpukan data. Sehingga sangat disayangkan jika tidak di analisa kembali. produk yang ditawarkan bermacam dengan berbagai macam produk, dan terkadang merk mempengaruhi masyarakat untuk membeli produk tersebut, untuk mengetahui produk dengan penjualan terbanyak dan keterkaitan produk satu dengan yang lainnya diperlukan salah satu algoritma yang ada di algoritma data mining yaitu algoritma apriori untuk dapat mengetahuinya, dan dengan bantuan aplikasi tanagra, produk yang muncul secara bersamaan dapat diketahui. Algoritma apriori termasuk jenis aturan asosiasi pada data

mining. Salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (frequent pattern mining). Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolak ukur, yaitu : support dan confidence. Support (nilai penunjang) adalah persentase kombinasi item tersebut dalam database, sedangkan confidence (nilai kepastian) adalah kuatnya hubungan antar-item dalam aturan asosiasi. Algoritma apriori dapat membantu untuk pengembangan strategi pemasaran.

Kata Kunci: Asosiasi, apriori, penjualan

PENDAHULUAN

Semakin banyaknya persaingan dalam dunia bisnis khususnya dalam industri penjualan menuntut pihak manajemen untuk putar otak mencari peluang untuk tetap berkecimpung dalam bisnisnya. Untuk menyiasati hal tersebut, tentunya pihak manajemen harus mampu menganalisis suatu data yang ada untuk di jadikan bahan acuan untuk menganalisa data tersebut.

Ketersediaan detail informasi transaksi pelanggan mendorong pengembangan teknik yang secara otomatis mencari hubungan antaritem dalam data di database. Sebagai contoh data didapat dari scanner bar-code di supermarket. Database penjualan menyimpan jumlah record transaksi penjualan yang sangat besar. Setiap record memberikan daftar item barang yang dibeli oleh pelanggan dalam satu transaksi. Manager mungkin akan tertarik untuk mengetahui jika beberapa kelompok item barang secara konsisten dibeli secara bersama. Manager dapat menggunakan data tersebut dalam pengaturan layout toko untuk meletakkan item barang secara optimal dengan keterkaitan satu dengan lainnya, dapat pula digunakan dalam promosi, atau dalam design katalog dan untuk mengidentifikasi segmen pelanggan berdasar pola pembelian.

Menuntut para pengembang untuk menemukan suatu strategi yang dapat meningkatkan penjualan dan pemasaran produk yang dijual, salah satunya adalah dengan pemanfaatan data penjualan produk. Dengan adanya kegiatan penjualan setiap hari, data semakin lama akan semakin bertambah banyak. Data tersebut tidak hanya berfungsi sebagai arsip bagi perusahaan, data tersebut dapat dimanfaatkan dan diolah menjadi informasi yang berguna untuk peningkatan penjualan dan promosi produk.

Proses untuk menguraikan penemuan pengetahuan di dalam database seperti yang dilakukan untuk melihat keterkaitan penjualan antar item diatas merupakan suatu konsep yang di sebut dengan data mining. Data mining merupakan konsep utama dalam *business intelligence* (BI) serta online *analytical processing* (OLAP)(Vercellis, 2009). Terdapat banyak algoritma data mining yang jika dapat diterapkan dalam proses bisnis, akan memberikan nilai positif bagi peningkatan kinerja proses bisnis tersebut yang berujung pada peningkatan keuntungan dari bisnistersebut.

Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada. Para pengambil keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil keputusan, hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah besar, yang disebut dengan data mining. Penggunaan teknik data mining diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi didalam gudang data sehingga menjadi informasi yang berharga.

Algoritma asosiasi merupakan suatu bentuk algoritma dalam data mining yang memberikan informasi hubungan antar item data didatabase. Algoritma tersebut dapat dimanfaatkan secara luas dalam proses bisnis diantaranya dalam proses penjualan. *Data mining* algoritma asosiasi dapat membantu dalam proses penjualan dengan memberikan hubungan antar data penjualan yang dilakukan pelanggan sehingga akan didapat pola pembelian pelanggan. Pebisnis dapat memanfaatkan informasi tersebut untuk mengambil tindakan bisnis yang sesuai.

Di dalam bidang usaha retail, metode *association rule mining* ini lebih dikenal dengan istilah analisa keranjang belanja atau *market*

basket analysis(Gunadi, 2013). Algoritma apriori adalah algoritma pengambilan data dengan aturan asosiatif (*Association rule*) untuk menentukan hubungan asosiatif suatu kombinasi item(Kusrini, 2007).

Anna Celluler merupakan salah satu toko yang bergerak dalam bidang penjualan perangkat *mobile* atau *smartphone* yang berada di daerah Jakarta. Berbagai macam jenis perangkat celluler ditawarkan di toko tersebut dengan berbagai macam variasi harga tergantung dari merk dan spesifikasi dari produk tersebut. Proses pencatatan transaksi yang dilakukan masih mengandalkan nota atau faktur untuk bukti transaksi, Data transaksi yang sudah dilakukan sampai pada proses pembuatan nota atau faktur kemudian di rekap oleh bagian administrasi untuk dilaporkan pada pemilik toko tersebut. Banyaknya data tersebut membutuhkan sebuah metode atau teknik yang dapat merubah banyaknya data agar dapat dimanfaatkan menjadi sebuah informasi berharga atau pengetahuan yang bermanfaat. Salah satu teknik yang dapat mengolah informasi dari kumpulan data yang besar menggunakan data mining.

Data mining adalah proses mengekstraksi informasi atau sesuatu yang penting atau menarik dari data yang ada di dalam database sehingga menghasilkan informasi yang sangat berharga. Data mining dapat mengolah informasi dari kumpulan data yang sangat besar dan dalam melakukan pencarian data membentuk pola yang biasa disebut dengan *association rule/Market basket analysis*(Kurniawati, 2014). *Market basket analysis* adalah suatu metode analisa atas perilaku konsumen secara spesifik dari suatu golongan/kelompok tertentu. *Market basket analysis* umumnya dimanfaatkan sebagai titik awal pencarian pengetahuan dari suatu transaksi data ketika kita tidak mengetahui pola spesifik apa yang kita cari(Gunadi, 2012). Teknik analisa keranjang pasar merupakan teknik yang mengadaptasi ilmu data mining. Teknik ini digunakan untuk merancang suatu strategi penjualan dan pemasaran barang melalui proses pencarian asosiasi atau hubungan antar item data dari suatu basis data relasional. Pencarian asosiasi berawal dari pengolahan data transaksi pembelian barang dari setiap pembeli, kemudian dicari hubungan antar barang-barang yang dibeli. Pencarian informasi ini hampir sama dengan mencari peluang kemunculan barang yang dibeli sesuai dengan kebiasaan berbelanja masyarakat dan jumlah transaksi yang ada.

Kebutuhan *market basket analysis* berawal dari keakuratan dan manfaat yang dihasilkannya dalam wujud aturan asosiasi (*association rules*). Yang dimaksud dengan *association rules* adalah pola-pola keterkaitan data dalam basis data. Data

ini yang nantinya akan di analisis seorang analisis data untuk menemukan pola dari data penjualan yang sudah terkumpul setiap hari guna untuk melakukan proses pengambilan keputusan. *Association Rule* yang dimaksud dilakukan melalui mekanisme penghitungan *support* dan *confidence* dari suatu hubungan item. Sebuah rule asosiasi dikatakan interesting jika nilai *support* adalah lebih besar dari minimum *support* dan juga nilai *confidence* adalah lebih besar dari minimum *confidence* (Nurcahyo, 2013). Proses pencarian asosiasi ini menggunakan algoritma apriori, yang berfungsi untuk membentuk kandidat kombinasi item yang mungkin, lalu diuji apakah kombinasi tersebut memenuhi parameter *support* dan *confidence* minimum yang merupakan nilai ambang yang diberikan oleh user.

Pada penelitian ini, penulis akan melakukan analisis terhadap data transaksi yang sudah dilakukan dengan mekanisme perhitungan nilai *support* dan *confident* dari suatu hubungan item, yang nantinya hasil nilai *support* dan *confident* yang ada akan digunakan untuk proses pengambilan keputusan oleh pihak manajemen untuk meningkatkan strategi pemasaran.

BAHAN DAN METODE

a. Data Mining

Data mining telah menarik banyak perhatian dalam dunia sistem informasi dan dalam masyarakat secara keseluruhan dalam beberapa tahun terakhir, karena ketersediaan luas dalam jumlah besar data dan kebutuhan segera untuk mengubah data tersebut menjadi informasi yang berguna dan pengetahuan (Witten, 2012). Informasi dan pengetahuan yang diperoleh dapat digunakan untuk aplikasi mulai dari pasar analisis, deteksi penipuan, dan retensi pelanggan, untuk pengendalian produksi dan ilmu pengetahuan eksplorasi (Han & Kamber, 2007). Adanya ketersediaan data yang melimpah, kebutuhan akan informasi atau pengetahuan sebagai sarana pendukung dalam pengambilan keputusan baik bagi individu, organisasi, perusahaan dan pemerintahan.

Banyaknya data, ditambah dengan kebutuhan untuk alat analisis data yang kuat, telah digambarkan sebagai kaya data tapi miskin informasi. Jumlah data yang tumbuh secara cepat, dikumpulkan dan disimpan dalam repositori data yang besar dan banyak, telah jauh melampaui kemampuan manusia untuk memahami data-data tersebut tanpa mampu mengelola data tersebut. Akibatnya, data yang dikumpulkan dalam repositori data yang besar menjadi "kuburan data" (Han & Kamber, 2007).

Hal ini melatarbelakangi lahirnya suatu cabang ilmu pengetahuan baru yaitu *data mining*. *Data mining* adalah untuk mengekstraksikan atau "menambang" pengetahuan dari kumpulan banyak data (Han & Kamber, 2007). *Data mining* adalah teknik yang merupakan gabungan metode-metode analisis data secara berkesinambungan dengan algoritma-algoritma untuk memproses data berukuran besar. *Data mining* merupakan proses menemukan informasi atau pola yang penting dalam basis data berukuran besar dan merupakan kegiatan untuk menemukan informasi atau pengetahuan yang berguna secara otomatis dari data yang jumlahnya besar (Kusrini, 2009). *Data mining*, sering juga disebut *knowledge discovery in database (KDD)*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan pola keteraturan, pola hubungan dalam set data berukuran besar. Keluaran dari data mining ini dapat dijadikan untuk memperbaiki pengambilan keputusan di masa depan. Dalam *data mining* data disimpan secara elektronik dan diolah secara otomatis, atau setidaknya disimpan dalam komputer. *Data mining* adalah tentang menyelesaikan masalah dengan menganalisa data yang telah ada dalam database (Kusrini, 2009).

Siklus hidup proyek *data mining* menurut *Cross-Industry Standard Proses for Data Mining (CRISP-DM)* yang dikembangkan tahun 1996 terbagi dalam 6 fase (Kusrini, 2009). Berikut gambar dari Siklus hidup proyek *data mining*

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
 - a. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
 - b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mining*.
 - c. Menyiapkan strategi awal untuk mencapai tujuan.
2. Fase pemahaman data (*Data Understanding Phase*)
 - a. Mengumpulkan data.
 - b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
 - c. Mengevaluasi kualitas data.
 - d. Jika diinginkan, pilih sebagian kecil group data yang mungkin mengundang pola dari permasalahan.
3. Fase pengolahan data (*Data Preparation Phase*)
 - a. Siapkan data awal, kumpulkan data yang akan digunakan untuk keseluruhan fase berikutnya. Fase ini merupakan pekerjaan

- berat yang perlu dilaksanakan secara intensif.
- b. Pilih kasus dan variabel yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
 - c. Lakukan perubahan pada beberapa variabel jika dibutuhkan.
 - d. Siapkan data awal sehingga siap untuk perangkat pemodelan.
4. Fase Pemodelan (*Modelling Phase*)
 - a. Pilih dan aplikasikan teknik pemodelan yang sesuai.
 - b. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan *data mining* yang sama.
 - c. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik *data mining* tertentu.
 5. Fase Evaluasi
 - a. Pengevaluasi satu atau lebih model yang digunakan dalam fase permodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
 - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
 - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
 - d. Mengambil keputusan yang berkaitan dengan penggunaan hasil dari *data mining*.
 6. Fase Penyebaran
 - a. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.
 - b. Contoh sederhana penyebaran: pembuatan laporan.
 - c. Contoh kompleks penyebaran: penerapan proses *data mining* secara paralel pada department lain.
- Menurut Larose dalam bukunya yang berjudul "*Discovering Knowledge in Data: An Introduction to Data Mining*", *datamining* dibagi menjadi beberapa kelompok berdasarkan tugas/pekerjaan yang dapat dilakukan (Larose, 2005), yaitu :
1. Deskripsi

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.
 2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan baris data (*record*) lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.
 3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.
 4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.
 5. Pengklasteran (*Clustering*)

Pengklasteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas obyek-obyek yang memiliki kemiripan. Klaster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan *record* dalam klaster yang lain. Berbeda dengan klasifikasi, pada pengklasteran tidak ada variabel target. Pengklasteran tidak melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target, akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.
 6. Asosiasi

Tugas asosiasi dalam *data mining* adalah untuk menemukan atribut yang muncul dalam satu waktu. Salah satu implementasi dari asosiasi adalah *market basket analysis* atau analisis keranjang belanja, sebagaimana yang akan dibahas dalam penelitian ini.

Dalam buku "Data Mining – Meramalkan Bisnis Perusahaan, Feri Sulianta & Dominikus Juju, 2010" menyatakan bahwa secara terperinci, ada empat tahap yang dilalui dalam *Data Mining* antara lain :

 1. Tahap pertama: *Precise statement of the problem* (mendefinisikan permasalahan yang ingin diketahui). Misalnya ingin mengetahui

apakah seorang customer berpotensi memiliki kredit macet, atau mengidentifikasi seorang customer apakah akan pindah ke kompetitor bisnis kita, dan lain sebagainya. Setelah menemukan pertanyaan bisnis yang perlu dijawab oleh data mining, selanjutnya tentukan tipe tugas dasar untuk menjawab pertanyaan bisnis tersebut. Tugas dasar yang menjadi dasar algoritma data mining adalah klasifikasi, regresi, segmentasi, asosiasi dan sequence analisis.

2. Tahap kedua: *Initial Exploration* (Mempersiapkan data yang menjadi sumber untuk *data mining* termasuk data "*cleaning*" untuk dipelajari polanya). Setelah menentukan definisi masalah, langkah berikutnya adalah mencari data yang mendukung definisi masalah. Menentukan porsi data yang digunakan men-training data mining berdasarkan algoritma data mining yang telah dibuat. Setelah persiapan data selesai dilakukan, langkah berikutnya adalah memberikan sebagian data kedalam algoritma data mining.
3. Tahap Ketiga: *Model building and validation*. Validasi apakah data mining memberikan prediksi yang akurat. Setelah *training* data selesai dilakukan, data mining tersebut perlu di-"uji" atau di-validasi keakuratannya terhadap data testing.
4. Tahap Ke-empat: *Deployment*. Tahap ini memilih aplikasi yang tepat terhadap *data mining* untuk membuat prediksi

b. Algoritma Apriori

Algoritma Apriori adalah suatu algoritma dasar yang diusulkan oleh Agrawal & Srikant pada tahun 1994 untuk penentuan *frequent itemsets* untuk aturan asosiasi *boolean*. Algoritma apriori termasuk jenis aturan asosiasi pada *data mining*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis*. Analisis asosiasi atau *association rule mining* adalah teknik data mining untuk menemukan aturan suatu kombinasi *item* (Larose, 2005). Salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (*frequent pattern mining*). Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolak ukur, yaitu: *support* dan *confidence*. *Support* (nilai penunjang) adalah persentase kombinasi *item* tersebut dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar-*item* dalam aturan asosiasi (Kusrini, 2009).

c. Analisis Pola Frekuensi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam basis data. Nilai *support* sebuah *item* diperoleh dengan menggunakan rumus berikut:

$$\text{support}(A) = \frac{\text{jumlah transaksi mengandung } A}{\text{total transaksi}}$$

Sementara, nilai *support* dari 2 *item* diperoleh dengan menggunakan rumus:

$$\text{support}(A, B) = \frac{P(A \cap B)}{\sum \text{transaksi mengandung } A \text{ dan } B} = \frac{\sum \text{transaksi}}{\sum \text{transaksi}}$$

Frequent itemset menunjukkan *itemset* yang memiliki frekuensi kemunculan lebih dari nilai minimum yang ditentukan (ϕ). Misalkan $\phi = 2$, maka semua *itemsets* yang frekuensinya kemunculannya lebih dari atau sama dengan 2 kali disebut *frequent*. Himpunan dari *frequent itemset* dilambangkan dengan F_k .

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif $A \rightarrow B$. Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dengan rumus berikut:

$$\text{confidence} = P(B|A) = \frac{\sum \text{transaksi mengandung } A \text{ dan } B}{\sum \text{transaksi mengandung } A}$$

Untuk menentukan aturan asosiasi yang akan dipilih maka harus diurutkan berdasarkan $\text{Support} \times \text{Confidence}$. Aturan diambil sebanyak n aturan yang memiliki hasil terbesar.

d. Tanagra

Tanagra adalah software Data Mining bebas untuk tujuan akademik dan penelitian ini mengusulkan beberapa metode data mining dari analisis eksplorasi data, pembelajaran statistik, pembelajaran mesin dan daerah database. Tanagra adalah "proyek *open source*" karena setiap peneliti dapat mengakses ke kode sumber, dan menambahkan algoritma sendiri, sejauh dia setuju dan sesuai dengan lisensi distribusi perangkat lunak.

Tujuan utama dari proyek Tanagra adalah memberikan peneliti dan mahasiswa yang mudah untuk menggunakan perangkat lunak data mining, sesuai dengan norma-norma yang hadir dari pengembangan perangkat lunak dalam domain ini (terutama dalam desain GUI dan cara untuk menggunakannya), dan memungkinkan untuk menganalisis baik data yang nyata atau sintesis.

Tujuan kedua Tanagra adalah untuk mengusulkan kepada peneliti arsitektur yang memungkinkan mereka untuk dengan mudah menambahkan metode penambangan data

mereka sendiri, untuk membandingkan kinerja mereka. Tanagra bertindak lebih sebagai platform eksperimental untuk membiarkan mereka pergi ke penting dari pekerjaan mereka, pengeluaran mereka untuk berurusan dengan bagian menyenangkan dalam *programmation* semacam ini alat pengelolaan data (Julsam, 2009).

Tujuan ketiga dan terakhir, arah pengembang pemula, terdiri dalam menyebarkan metodologi yang mungkin untuk membangun perangkat lunak semacam ini. Mereka harus mengambil keuntungan dari akses gratis ke kode sumber, untuk melihat bagaimana perangkat lunak semacam ini dibangun, masalah untuk menghindari, langkah-langkah utama dari proyek ini, dan alat-alat dan perpustakaan kode yang digunakan untuk. Dengan cara ini, Tanagra dapat dianggap sebagai alat pedagogis untuk belajar teknik pemrograman

Metode Penelitian

Penelitian adalah mencari melalui proses yang metodis untuk menambahkan pengetahuan itu sendiri dan dengan yang lainnya, oleh penemuan fakta dan wawasan tidak biasa. Pengertian lainnya, penelitian adalah sebuah kegiatan yang bertujuan untuk membuat kontribusi orisinal terhadap ilmu pengetahuan. Penelitian ini adalah penelitian eksperimen dengan metode penelitian sebagai berikut:

1. Pengumpulan data

Pada pengumpulan data dijelaskan tentang bagaimana dan dari mana data dalam penelitian ini didapatkan, ada dua tipe dalam pengumpulan data, yaitu pengumpulan data primer dan pengumpulan data sekunder. Data primer adalah data yang dikumpulkan pertama kali untuk melihat apa yang sesungguhnya terjadi. Data sekunder adalah data yang sebelumnya pernah dibuat oleh seseorang baik di terbitkan atau tidak (Vercelis, 2009). Dalam pengumpulan data primer dalam penelitian ini menggunakan metode observasi dan interview, dengan menggunakan data-data yang berhubungan dengan penjualan produk tahun 2015. Data penjualan pada Anna Celluler selama ini tidak tersusun dengan baik, sehingga data penjualan yang semakin hari semakin banyak tersebut hanya berfungsi sebagai arsip bagi perusahaan dan tidak dapat dimanfaatkan perusahaan untuk pengembangan strategi seperti pemasaran atau yang lainnya. Oleh sebab itu perusahaan memerlukan sistem untuk mengolah data yang dapat menghasilkan data penjualan produk yang paling sering dibeli, sehingga produk yang paling sering dibeli tersebut dapat menjadi acuan untuk

mengembangkan strategi pemasaran produk tersebut kepada konsumen.

2. Pengolahan awal data

Jumlah data awal yang diperoleh dari pengumpulan data dalam 3 bulan yaitu bulan Juli sd bulan September 2015 terdiri dari banyak produk dan jumlahnya sangat besar, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*). Untuk mendapatkan datayang berkualitas, beberapa teknik yang dilakukan yaitu (Kothari, 2004) : data validation, data integration and transformation dan data *size reduction and discritization*. Sehingga perlu dilakukan filtering terhadap produk yang akan dilakukan analisis. Dalam hal ini penulis mengelompokkan barang yang diperjual belikan selama kurun waktu 3 bulan yang dimulai dari Juli 2015 sd September 2015.

1. Model yang diusulkan

Model yang diusulkan pada penelitian ini berdasarkan *state of the art* tentang prediksi hasil pemilihan umum adalah dengan menerapkan algoritma apriori untuk melakukan analisis terhadap produk yang sudah diperjualbelikan dan melihat aturan asosiasi dari data yang sudah ada.

2. Eksperiment dan pengujian model

Untuk melakukan eksperiment dan pengujian model, dilakukan pengelompokan data dan disesuaikan dengan data yang ada di algoritma apriori. Kemudian data di hitung nilai support mulai dari perhitungan itemset pertama, itemset kedua dan seterusnya. Selanjutnya menghitung nilai confident dari data yang ada untuk mengetahui asosiasi dari data yang ada.

HASIL DAN PEMBAHASAN

Prosedur Sistem berjalan dalam penelitian ini yaitu Pelanggan datang ke toko kemudian memilih barang yang tersedia berdasarkan yang ada di *display* toko, jika tidak ada yang cocok maka pelanggan akan memilih barang yang lain atau meninggalkan toko.

A. Perhitungan Algoritma Apriori

Berdasarkan transaksi penjualan produk pada toko celluler, transaksi tersebut dapat diakumulasikan. Akumulasi transaksi penjualan produk diperoleh dari penjualan mingguan yang diambil dari 3 teratas laporan mingguan, dapat dilihat dalam tabel dibawah ini :

Tabel 1. Pola Transaksi Penjualan smartphone

No.	Itemset
1	Samsung, Advance, Lenovo
2	Samsung, Apple, Lenovo
3	Huawei, Advance, Lenovo
4	Samsung, Apple, Advance
5	Samsung, Apple, Lenovo
6	Samsung, Huawei, Lenovo
7	Samsung, Apple, Lenovo
8	Samsung, Apple, Huawei
9	Apple, Huawei, Advance
10	Samsung, Huawei, Advance
11	Samsung, Advance, Lenovo
12	Apple, Huawei, Advance

Sumber : Anna Celluler (2016)

Format tabular data transaksi mingguan, bila dibentuk akan tampak seperti tabel berikut ini:

Tabel 2. Tabel Format Tabular Data Transaksi

Minggu	Samsu ng	App le	Huaw ei	Advan ce	Leno vo
1	1	0	0	1	1
2	1	1	0	0	1
3	0	0	1	1	1
4	1	1	0	1	0
5	1	0	0	1	1
6	1	0	1	0	1
7	1	1	0	0	1
8	1	1	1	0	0
9	0	1	1	1	0
10	1	0	1	1	0
11	1	0	0	1	1
12	0	1	1	1	0

Sumber : (Penelitian tahun 2016)

1. Analisis Pola Frekuensi Tinggi

Berikut ini adalah penyelesaian dengan contoh k1 berdasarkan data yang sudah disediakan pada tabel berikut ini.

Proses pembentukan C_1 atau disebut dengan 1 itemset dengan jumlah minimum support = 30%. Berikut merupakan perhitungan pembentukan 1 itemset:

$$S(\text{samsung}) = \frac{\sum \text{transaksi yang mengandung samsung}}{\sum 12} = \frac{9}{12} * 100\% = 75\%$$

$$S(\text{apple}) = \frac{\sum \text{transaksi yang mengandung apple}}{\sum 12} = \frac{6}{12} * 100\% = 50\%$$

$$S(\text{huawei}) = \frac{\sum \text{transaksi mengandung huawei}}{\sum 12} = \frac{6}{12} * 100\% = 50\%$$

$$S(\text{advance}) = \frac{\sum \text{transaksi mengandung advance}}{\sum 12} = \frac{8}{12} * 100\% = 66.66\%$$

$$S(\text{lenovo}) = \frac{\sum \text{transaksi mengandung Lenovo}}{\sum 12} = \frac{7}{12} * 100\% = 58.33\%$$

Tabel 3. Support dari tiap item

Itemset	Support
Samsung	75 %
Apple	50 %
Huawei	50 %
Advance	66.66 %
Lenovo	58.33 %

Sumber : Hasil Penelitian (2016)

Proses selanjutnya adalah pembentukan C_2 atau disebut dengan 2 itemset dengan jumlah minimum support = 30%. Berikut merupakan perhitungan pembentukan C_2 atau 2 itemset:

$$S(\text{Samsung,apple}) = \frac{\sum \text{transaksi yang mengandung samsung dan apple}}{\sum 12} * 100\% = \frac{4}{12} * 100\% = 33.33\%$$

$$S(\text{Samsung,huawei}) = \frac{\sum \text{transaksi yang mengandung samsung dan huawei}}{\sum 12} * 100\% = \frac{2}{12} * 100\% = 16.66\%$$

$$S(\text{Samsung,Advance}) = \frac{\sum \text{transaksi yang mengandung samsung dan advance}}{\sum 12} * 100\% = \frac{5}{12} * 100\% = 41.66\%$$

$$S(\text{samsung,lenovo}) = \frac{\sum \text{transaksi yang mengandung samsung dan lenovo}}{\sum 12} * 100\% = \frac{6}{12} * 100\% = 50\%$$

$$S(\text{apple, huawei}) = \frac{\sum \text{transaksi yang mengandung apple dan huawei}}{\sum 12} * 100\% = \frac{3}{12} * 100\% = 25\%$$

$$S(\text{apple,advance}) = \frac{\sum \text{transaksi yang mengandung apple dan advance}}{\sum 12} * 100\% = \frac{3}{12} * 100\% = 25\%$$

$$S(\text{apple, lenovo}) = \frac{\sum \text{transaksi yang mengandung apple dan lenovo}}{\sum 12} * 100\% = \frac{2}{12} * 100\% = 16.66\%$$

$$S(\text{huawei, advance}) = \frac{\sum \text{transaksi yang mengandung huawei dan advance}}{\sum 12} * 100\% = \frac{4}{12} * 100\% = 33.33\%$$

$$S(\text{huawei,lenovo}) = \frac{\sum \text{transaksi yang mengandung huawei dan lenovo}}{\sum 12} * 100\% = \frac{2}{12} * 100\% = 16.66\%$$

$$S(\text{advance,lenovo}) = \frac{\sum \text{transaksi yang mengandung advance dan lenovo}}{\sum 12} * 100\% = \frac{4}{12} * 100\% = 25\%$$

Tabel 4. Calon 2-itemset

Itemset	Jumlah	Support
Samsung, Apple	4	33.33 %
Samsung, Huawei	2	16.66 %
Samsung, Advance	5	41.66 %
Samsung, Lenovo	6	50 %
Apple, Huawei	3	25 %
Apple, Advance	3	25 %

Apple, Lenovo	2	16.66 %
Huawei, Advance	4	33.33 %
Huawei, Lenovo	2	16.66 %
Advance, Lenovo	4	33.33 %

Sumber : Hasil Penelitian (2016)

Minimal support yang ditentukan adalah 30 %, jadi kombinasi 2 itemset yang tidak memenuhi minimal support akan dihilangkan, terlihat seperti table dibawah ini:

Tabel 5. Minimal Support 2 itemset 35%

Itemset	Support
Samsung, Apple	33.33 %
Samsung, Advance	41.66 %
Samsung, Lenovo	50 %
Huawei, Advance	33.33 %
Advance, Lenovo	33.33 %

Sumber : Hasil Penelitian (2016)

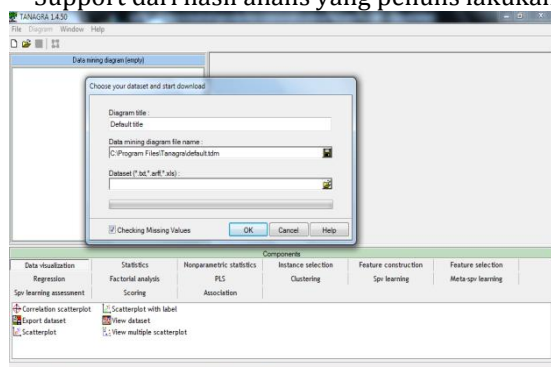
B. Perhitungan Algoritma Apriori dengan Tanagra

Algoritma Apriori pada Tanagra dapat terbentuk dengan algoritma atau langkah-langkah yang telah ditentukan.

Algoritma ini terdiri dari dua algoritma yaitu:

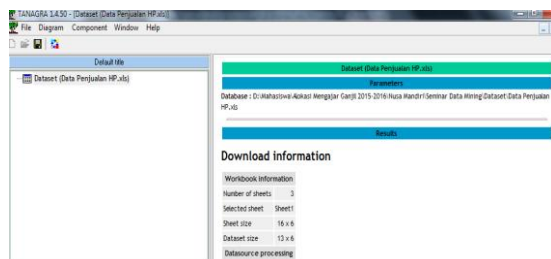
1. Algoritma Support

Algoritma penentuan Support dapat dilihat pada Algoritma dibawah ini yang terdiri dari input, Output dan proses. Berikut hasil Support dari hasil analisis yang penulis lakukan.



Sumber : Hasil Penelitian (2016)

Gambar 1. Proses pengambilan data



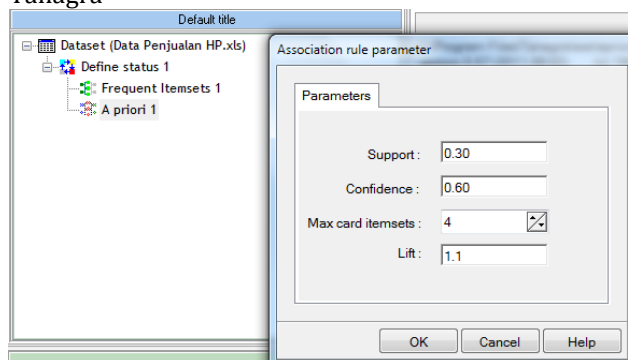
ITEMSETS [#5 itemsets loaded]		
N°	Description	Support
1	Apple ^ Samsung	33.3
2	Huawei ^ Advance	33.3
3	Lenovo ^ Advance	33.3
4	Lenovo ^ Samsung	50.0
5	Advance ^ Samsung	41.7

Sumber : Hasil Penelitian (2016)

Gambar 2 . tampilan support

2. Algoritma Confidence

Algoritma penentuan Confidence dapat dilihat pada Algoritma dibawah ini yang terdiri dari input, Output dan proses. Berikut algoritma dari algoritma confident *If support ≥ 60% then min.confidence terpenuhi Else eliminasi*. Dan berikut hasil dari penerapan menggunakan Tanagra



Gambar 3. Algoritma Confident

RULES

Number of rules : 2					
N°	Antecedent	Consequent	Lift	Support (%)	Confidence (%)
1	"Samsung=true"	"Lenovo=true"	1.14286	50.000	66.667
2	"Lenovo=true"	"Samsung=true"	1.14286	50.000	85.714

Computation time : 172 ms.
Created at 11/12/2015 8:45:17 PM

Sumber : Hasil Penelitian (2016)

Gambar 4. Hasil Algoritma Confident

PENUTUP

Algoritma Apriori dapat diterapkan untuk mendukung strategi marketing di suatu perusahaan atau institusi lainnya. Informasi yang berkaitan dengan pelaksanaan promosi dapat tersedia dengan cepat, sehingga pihak manajemen dapat melakukan pengambilan keputusan dengan cepat. Pelaksanaan strategi promosi pada perusahaan sangat dibantu dengan adanya penerapan algoritma Apriori ini sehingga diharapkan efektifitas pelaksanaan strategi akan dapat semakin ditingkatkan. Penerapan algoritma Apriori yang dilakukan melalui perangkat lunak yang dirancang terbukti menunjukkan hasil yang sama dibandingkan dengan melakukan perhitungan secara manual ataupun dengan perangkat lunak sejenis. Hal ini dibuktikan melalui perhitungan nilai support dan

confidence yang menunjukkan hasil yang sama. Adapun saran dari hasil penelitian adalah sebagai berikut : algoritma apriori ini dapat diterapkan dalam proses pelaksanaan strategi dan proses lainnya yang dapat melibatkan hubungan antar beberapa item dalam lingkungan lainnya tidak hanya terbatas pada perusahaan saja seperti misalnya pada strategi *cross market analysis*. Selanjutnya penulis juga menyarankan agar dapat membandingkan metode pengambilan keputusan dengan menggunakan algoritma apriori ini dengan teknik yang lainnya

REFERENSI

- Dawson, C. W. (2009). *Projects in Computing and Information System A Student's Guide*. England: Addison-Wesley.
- Gunadi - Penerapan Metode Data Mining Market Basket Analysis terhadap Penjualan Buku dengan Algoritma Apriori – 2012
- Han, J., & Kamber, M. , 2007, *Data Mining Concepts and Technique*. Morgan Kaufmann publisher
- Julsam dkk, Penambangan Data Format Text Excel dengan Software Tanagra, 2009.
- Kothari, C. R. (2004). *Research Methology methodes and Technique*. India: New Age Interntional.
- Kusrini, dan Luthfi, E. T. (2007). *Algoritma Data mining*. Yogyakarta: Andi.
- Kusrini, dan Luthfi, E. T. (2009). *Algoritma Data mining*. Yogyakarta: Andi.
- Kurniawati - Pemetaan Pola Hubungan Program Studi dengan Algoritma Apriori – 2014.
- Larose, D. T. (2005). *Discovering Knowledge in Data*. Canada: Wiley Interscience.
- Nurcahyo - Penerapan Data Mining dengan Algoritma Apriori untuk Mendukung Strategi Promosi Pendidikan
- Vercellis, C. (2009). *Business Intelligence : Data Mining and Optimization for Decision Making*. John Wiley dan Sons, Ltd
- Witten, H. I., Eibe, F., dan Hall, A. M. (2011). *Data Mining Machine Learning Tools and Techiques*. Burlington: Morgan Kaufmann Publisher.

BIODATA PENULIS



Mohammad Badrul, M.Kom

Penulis adalah Dosen Tetap di STMIK Nusa Mandiri Jakarta. Penulis Kelahiran di Bangkalan 01 Januari 1984. Penulis menyelesaikan Program Studi Strata 1 (S1) di Kampus STMIK Nusa Mandiri Prodi Sistem Informasi dengan gelar S.Kom pada tahun 2009 dan menyelesaikan progarm Srata 2 (S2) di Kampus yang sama dengan Prodi ilmu Komputer dengan gelar M.Kom pada tahun 2012. Selain mengajar, Penulis juga aktif dalam membimbing mahasiswa yang sedang melakukan penelitian khususnya di tingkat Strata 1 dan penulis juga terlibat dalam tim konsorsium di Jurusan Teknik Informatika STMIK Nusa Mandiri untuk penyusunan bahan ajar. Saat ini penulis memiliki Jabatan Fungsional Asisten ahli di kampus STMIK Nusa Mandiri Jakarta. Penulis tertarik dalam bidang kelimuan Data mining, Jaringan komputer, Operating sistem khususnya *open source, Database, Software engineering* dan *Research Metode*.